Lesson 7.2 Frequency Distributions

Objectives

- Organize data into a frequency distribution.
- Find the mean using a frequency distribution.
- Create a histogram from a frequency distribution.

In Lesson 7.1, you used a data table to organize production data for three workers. Make a frequency distribution to organize the data.

Skateboard Production Data										
Worker	**M**	**T**	**W**	**Th**	**F**	**M**	**T**	**W**	**Th**	**F**
A	1	2	2	3	3	4	5	4	5	5
B	6	1	2	5	3	2	3	2	7	1
C	7	6	5	4	2	3	2	3	2	2

Organizing Data

A **frequency distribution** which displays the number of times values occur in a set of data, is another useful tool for organizing information.

Example 1 Making a Frequency Distribution

Make a frequency distribution for Worker A.

Solution

Start by making a table with three columns, as shown below. List all the production levels for Worker A in the first column. For example, the first row labeled 5 means a worker produces 5 skateboards in one day.

Next, tally (count) the number of times Worker A meets each production level. For example, on the first Monday Worker A produced 1 skateboard. Therefore, place a tally mark in the row labeled 1. Continue through the rest of the day's list, placing tally marks in the appropriate rows. Finally, determine the class frequency by counting the number of tallies in each row.

Production Data for Worker A		
Skateboards	**Tally**	**Frequency**
5	///	3
4	//	2
3	//	2
2	//	2
1	/	1

The formula for finding the mean of a frequency distribution is written as follows.

Formula for the Mean

Let c represent the class, f represent the frequency, and n represent the number of tallies. To calculate the mean $\overline{x}$ from a frequency distribution table, use this formula

$$\overline{x} = \frac{\Sigma(c \bullet f)}{n}.$$

Example 2 Finding the Mean

Use the formula to find the mean of Worker A's production.

Solution

The mean, or average, number of skateboards made per day is the total number of skateboards made divided by the number of days. You can find the mean from the frequency distribution table. First multiply each skateboard number by its frequency. If you let c represent the number of skateboards and f represent the frequency, this product is $c \bullet f$.

Next, add each of the products. The Greek symbol Σ (sigma) means to add a set of numbers. Thus, the second step is to find

$$\Sigma(c \bullet f).$$

Finally, divide the sum by the total number n of tallies. The result is the mean and is usually denoted by $\overline{x}$.

Skateboards	Tally	Frequency	$(c \bullet f)$
5	///	3	15
4	//	2	8
3	//	2	6
2	//	2	4
1	/	1	1
		10	34
		$n = 10$	$\Sigma(c \bullet f) = 34$

Thus, $\overline{x} = \frac{34}{10} = 3.4$.

Critical Thinking How can you find the mode and the median by using the tally marks?

Ongoing Assessment

Make a frequency distribution table for Worker B and Worker C. Use the formula to find the mean for each worker. Use the tally marks to find the mode and median for each worker.

Example 3 Using Data from a Frequency Table

Use the frequency distribution table in Example 2.

a. On how many days did Worker A make fewer than three skateboards?

b. On what percent of the days did Worker A make more than the mean of the distribution?

Solution

Examine the frequency distribution table. The number in the frequency column is the number of days the worker equaled the production level in the first column.

a. Worker A made 1 skateboard on one day and 2 skateboards on two days. Thus, Worker A made fewer than 3 skateboards on three days.

b. The mean is 3.4. Worker A made 4 skateboards on two days and 5 skateboards on three days. Thus, Worker A made more than the mean on five out of ten days, or 50% of the days.

Histograms

A graph of the frequency distribution is a very useful means of displaying large sets of data. Consider the weights (in pounds) of thirty football players.

286, 234, 211, 268, 227,
273, 276, 250, 184, 230,
228, 202, 260, 205, 193,
250, 248, 234, 234, 197,
246, 224, 218, 235, 235,
253, 219, 241, 226, 246

Activity Creating a Histogram

Use the football players' weights given on the previous page to complete the activity below.

1. Identify the greatest and least weights in the group, 286 and 184. The difference between these numbers is the **range** of the data. What is the range for this data?

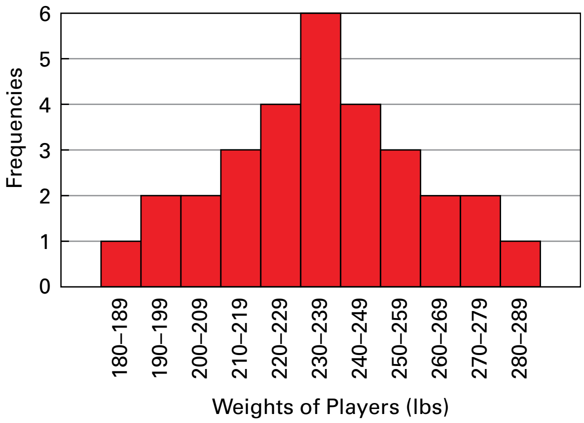
2. Group the data into weight intervals that each contain ten pounds. Complete the table on your paper.

Weight	Tally	Frequency
280–289	/	1
270–279	//	2

3. Draw a positive *x*-axis and a positive *y*-axis. Label the *x*-axis as "Weights of Players." The values on the *x*-axis begin with the weight interval 180–189 and end with 280–289.

4. Label the *y*-axis as "Frequencies." The frequencies begin with zero and go up to 6.

5. For each weight interval, draw a rectangle to show the frequency. Because each weight interval has equal value, the widths of the rectangles are the same.



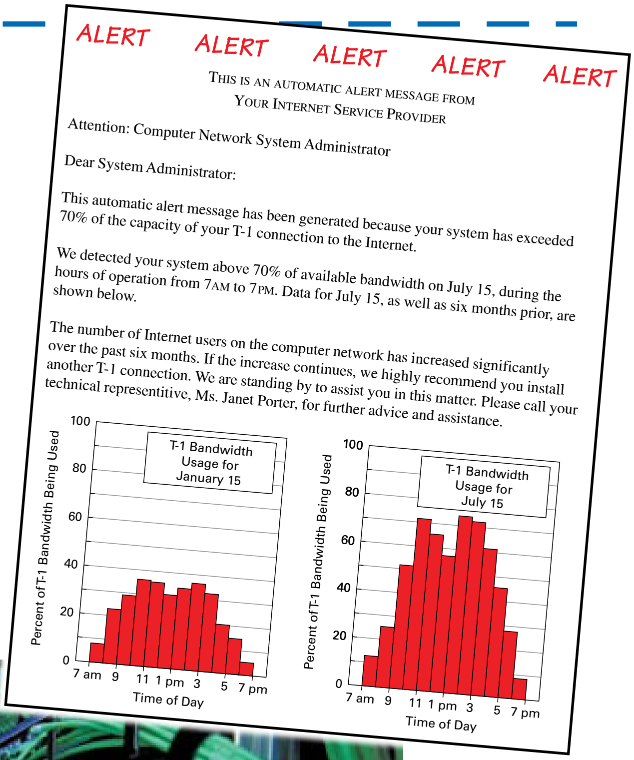
6. The height of each rectangle represents the frequency. The tallest rectangle represents the weight interval that occurs most often. Which weight interval is the mode for the weights of the football players?

Ongoing Assessment

Use your calculator to find the mean in the Activity. The graphing calculator uses a menu key with the abbreviation VARS. You can use the variable statistics key to find the mean ($\bar{x}$), the sum (Σ), and the number of data values (n).

WORKPLACE COMMUNICATION

For the July 15 data, which time interval represents the mode for bandwidth usage? If the rate of increase in the use of the Internet continues, how long do you think it will be until the company exceeds the capacity (100% of available bandwidth) of its T-1 connection? Why is there a "dip" in the middle of the histograms?



ALERT ALERT ALERT ALERT ALERT

THIS IS AN AUTOMATIC ALERT MESSAGE FROM
YOUR INTERNET SERVICE PROVIDER

Attention: Computer Network System Administrator

Dear System Administrator:

This automatic alert message has been generated because your system has exceeded 70% of the capacity of your T-1 connection to the Internet.

We detected your system above 70% of available bandwidth on July 15, during the hours of operation from 7AM to 7PM. Data for July 15, as well as six months prior, are shown below.

The number of Internet users on the computer network has increased significantly over the past six months. If the increase continues, we highly recommend you install another T-1 connection. We are standing by to assist you in this matter. Please call your technical representitive, Ms. Janet Porter, for further advice and assistance.

Lesson Assessment

Think and Discuss

1. What is a frequency distribution?

2. How is the expression $\Sigma(c \cdot f)$ used?

3. Explain how to find the mean of a frequency distribution.

4. Explain how to find the mode of a frequency distribution from a histogram.

Practice and Problem Solving

5. In one month a landscape company trimmed trees on 20 different days. The supervisor recorded the number of trees trimmed each day.

	M	Tu	W	Th	F
Week 1	10	30	7	5	25
Week 2	24	28	11	17	16
Week 3	19	28	21	26	35
Week 4	3	4	21	13	9

   a. Use the data to make a frequency distribution table.

   b. Make a histogram of the data.

   c. What is the range of the data?

   d. What is the mode of the data?

   e. What is the median of the data?

   f. What is the mean of the data?

6. A company manufactures linen goods. In one week, the company packed and shipped the following number of linen packages to 25 retail customers.

   80, 95, 50, 85, 100, 50, 75, 75, 80, 85, 70, 90, 95, 80, 85, 75, 80, 85, 70, 80, 100, 85, 60, 80, 95

   a. Set up a frequency table and tally the number of packages.

   b. Total the frequencies.

   c. Make a histogram.

   d. What is the mode number of packages sent?

   e. What is the median number of packages sent?

   f. What is the mean number of packages sent?

7. Use the heights in inches of the students in your class to make a table.

a. Make a frequency distribution for the data.

b. Use the table to make a histogram.

c. What is the range of the data?

d. What is the mode of the data?

e. What is the median of the data?

f. What is the mean of the data?

g. Which measure of central tendency best describes the height of the students in your class?

h. Are there any outliers in your data? If so, what effect does the outlier have on the mean of your data?

Mixed Review

Solve an equation for each problem.

8. Todd receives a commission of 3% on all sales over $5,000. What is Todd's commission on sales of $8,500?

9. A local survey shows that 7 out of 12 people wear passenger seat belts. How many people out of 216 passengers would you expect to wear seat belts?

10. The equation $y = 40x + 100$ models the cost in dollars, y, for renting a backhoe for a number of hours, x.

a. What is the slope of the equation?

b. What is the y-intercept?

c. Graph the equation.

d. What is the value of y when x is 2.5?

Lesson 7.3

Scatter Plots, Correlation, and Lines of Best Fit

Objectives

- **Interpret a scatter plot.**
- **Identify the correlation of data from a scatter plot.**
- **Find the line of best fit for a set of data.**

A video game company has recently noticed an increase in the number of defective video games produced. The production manager has asked a quality control technician the reason for this increase. The technician has a hunch that the increase is related to the absentee rate of the workers. She gathers the following data.

	Defective Video Games Test 1									
	M	T	W	Th	F	M	T	W	Th	F
Absentee Workers	9	11	5	4	2	7	7	11	10	5
Defective Video Games	9	10	11	6	3	6	8	9	7	4

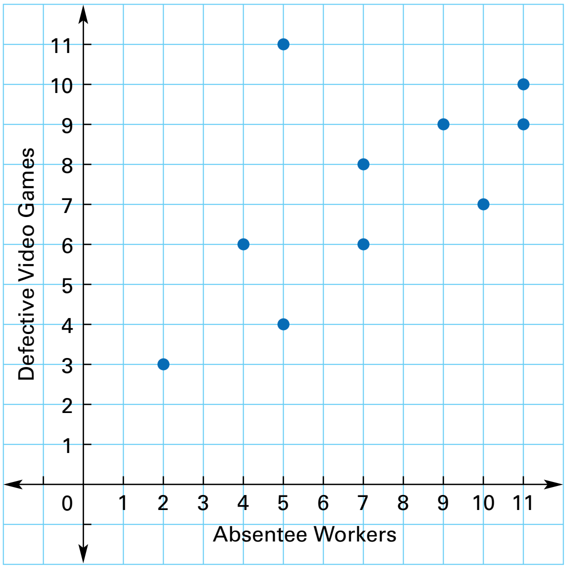
A **scatter plot** is a graph consisting of data points.

Example 1 Making a Scatter Plot

Make a scatter plot of the data from the opening paragraph of this lesson.

Solution

To check the technician's hunch, plot the data as individual points on a coordinate graph.



Activity Information from a Scatter Plot

1 Look at the days that have a low number of absentee workers. Do these days also have a low rate of defective video games?

2 Describe how the points are scattered. Do they appear to cluster around a straight line?

3 Does the line rise or fall as you move from left to right?

4 How does the line represent the relationship between the number of absentee workers and defective video games?

5 Create a scatter plot based on the data below. Complete Steps 1–4 with the following for Test 2.

	Defective Video Games Test 2									
	M	**T**	**W**	**Th**	**F**	**M**	**T**	**W**	**Th**	**F**
Absentee Workers	7	5	4	5	7	3	7	9	6	2
Defective Video Games	6	4	6	3	4	10	2	3	7	9

6 Create a scatter plot based on the data below. Complete Steps 1–4 with the following for Test 3.

	Defective Video Games Test 3									
	M	**T**	**W**	**Th**	**F**	**M**	**T**	**W**	**Th**	**F**
Absentee Workers	8	6	8	10	4	3	10	6	5	9
Defective Video Games	6	3	9	8	4	10	4	5	7	5

Correlation

If the slope is positive, a **positive correlation** exists; that is, as one variable increases, the other increases. If the slope is negative, a **negative correlation** exists; that is, as one variable increases, the other decreases. Sometimes there is a positive or negative correlation between data sets. Then you can predict the behavior or trend of one variable if you know the behavior of the other. If the plot is scattered in such a way that it does not approximate a line, there is **no correlation** between the sets of data.

Example 2 Interpreting Correlations

Describe the correlation of the data in Test 1 (in Example 1).

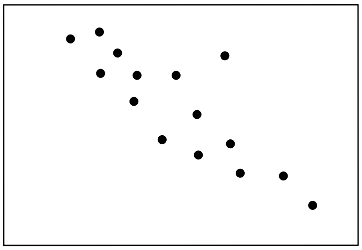
Solution

That data show a pattern that resembles a line rising from left to right. Therefore, the slope is positive, so there is a positive correlation. As the number of absentee workers increases, so does the number of defective video games.

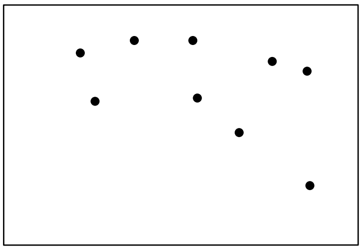
Example 3 Interpreting Correlations

Determine if the data in each scatter plot show a positive, negative, or no correlation. Describe each correlation as strong or weak.

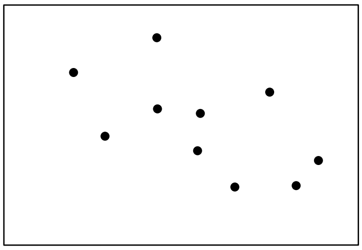
a.



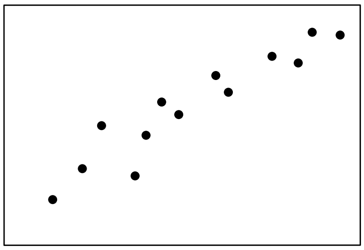
b.



c.



d.



Solution

a. Data are closely clustered from the upper left sloping downward to the lower right. These data have a strong negative correlation.

b. Data are scattered and show no relationship. These data are not correlated.

c. Data are somewhat clustered from the upper left sloping downward to the lower right. These data have a weak negative correlation.

d. Data are closely clustered from the lower left sloping upward to the upper right. These data have a strong positive correlation.

Ongoing Assessment

Describe the correlation in Test 2 and Test 3 in the Activity.

Line of Best Fit

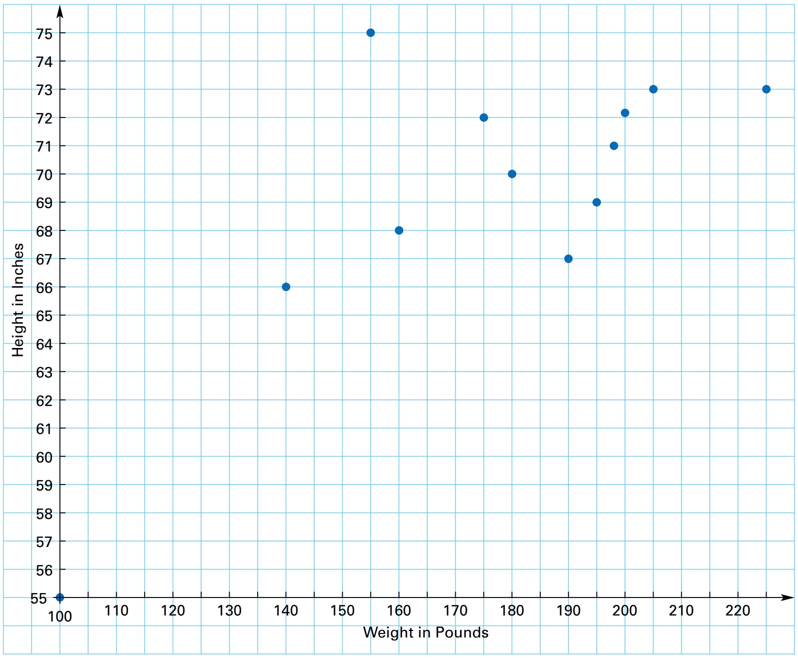
A **line of best fit** is a straight line that best represents the data on the scatter plot. This line may pass through some of the points, none of the points, or all of the points.

You can use a measure called the **correlation coefficient**, represented by the variable r, to describe how close the points cluster around the line of best fit. If all the points are on a line and the line has a positive slope, the correlation coefficient r is 1. If all the points are on a line and the line has a negative slope, r is -1. If the data are scattered at random, r is close to zero.

If r is close to 1 or -1, the trend of the data is well represented by a line, and there is a strong relationship between the data and the line. If r is close to zero, there is a weak relationship between the data and the line. In the real world, few relationships show a perfect correlation of 1 or -1.

Example 4 Naming a Line of Best Fit

The scatter plot below shows the relationship of weight (x-axis) to height (y-axis) of the first 12 students to try out for the varsity basketball team. You can use a graphing calculator to plot the data points. In the statistics menu of the graphing calculator, the line of best fit is called the **regression line**, LinReg $(ax + b)$. The correlation coefficient from the calculator is about 0.75. Determine a line of best fit. Write the equation. Round the slope and y-intercept to the nearest hundredth.



Solution

On your graphing calculator use the weights of the players for the data in L1.

100, 140, 155, 160, 175, 180, 190, 195, 198, 200, 205, 225

Enter the heights of the players for the data in L2. Be sure that you enter the players' heights in the same order as you did their weights.

55, 66, 75, 68, 72, 70, 67, 69, 71, 72.2, 73, 73

To turn on the DiagnosticOn feature, press 2nd CATALOG. Then scroll down to select DiagnosticOn. Press ENTER twice.

Press STAT. Select CALC and then LinReg($ax + b$). Press ENTER twice.

The slope is the value a, or 0.11. The y-intercept is the value b or 48.8. The equation is $y = 0.11x + 48.8$. You can see both the scatterplot and the line of best fit by entering the equation in the $Y=$ screen and graphing both Y_1 and Plot1 simultaneously.

Critical Thinking Explain what the correlation coefficient of 0.75 means with respect to the basketball players.

Example 5 Finding the Correlation Coefficient

Find the correlation coefficient for the points scored in the first ten games of a varsity basketball team's season. Graph the line of best fit using your graphing calculator.

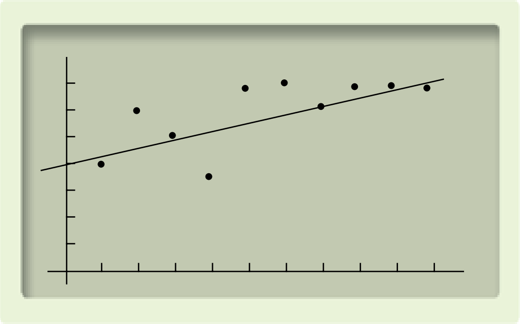
Game 1	Game 2	Game 3	Game 4	Game 5
36	58	49	27	67
Game 6	Game 7	Game 8	Game 9	Game 10
70	62	72	72	68

Solution

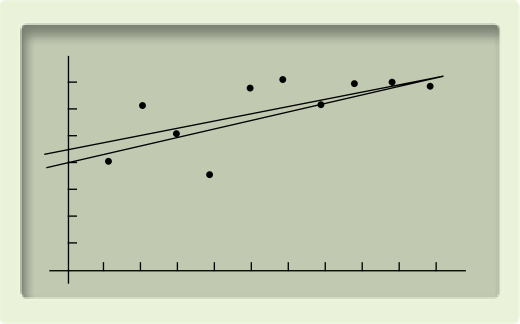
Enter the game numbers into L1 and the scores into L2. Be sure that Plot1 is turned on, scatter plot is the chosen type, and the Xlist is L1 and the Ylist is L2. If you want to see the scatter plot, be sure to set an appropriate viewing window. Be sure you have the DiagnosticOn feature turned on.

To find the correlation coefficient, press STAT. Select CALC and then LinReg($ax + b$). When you are returned to the home screen, press ENTER. The correlation coefficient is the value assigned to r. The correlation coefficient is about 0.71.

To graph the line of best fit over your scatter plot, you need to store the equation in Y_1. You can either round the correlation coefficient or assign the LinReg($ax + b$) to Y_1. Your graphing calculator screen should look similar to the screen shown below.



If your data set has an outlier, it may be better to use the **median-fit method** to determine the line of best fit. Press STAT and select CALC. Choose Med-Med. This method calculates the line of best fit excluding any outliers. If you graph the LinReg($ax + b$) and the Med-Med($ax + b$) over your scatter plot from Example 5 your graphing calculator screen should look similar to the one shown below.



Lesson Assessment

Think and Discuss

1. How are data displayed using a scatter plot?
2. Compare the meaning of a positive correlation and a negative correlation between two sets of data.
3. Explain how a “line of best fit” is used with a scatter plot.

Practice and Problem Solving

4. Jenny is measuring the relationship between the amount of physical exercise per week and age. She records the data as ordered pairs. The first number in each ordered pair is a person's age and the second number is the number of hours of physical activity per week for that person.

(20, 15), (22, 11), (30, 6), (30, 7), (34, 6.1), (26, 13), (26, 8.5), (18, 16), (36, 3), (36, 5.8), (28, 11), (30, 9), (40, 3)

a. Plot the points on a scatter plot.

b. Does the data show a correlation between age and the hours exercised each week? If so, what type of correlation?

c. Draw a line of best fit.

d. Use the slope and a point on the line to write an equation to approximate the line of best fit.

e. Enter the data in a graphing calculator and find the regression line.

f. Compare your line and the graphing calculator line and explain any differences.

5. The relationship between the number of hours studied and the score on an exam is represented by the ordered pairs below. The first number of the pair represents the number of hours studied and the second represents the test score.

(10, 100), (0, 50), (1, 60), (3, 70), (5, 75), (4, 80), (7, 95), (9, 80), (9, 90), (7, 70), (9, 80), (5, 85), (10, 85), (4, 70), (6, 85), (6, 90), (9, 80), (10, 95), (3, 75)

a. Draw a scatter plot for the data.

b. Find the mode of the test scores.

c. What was the score of the student who did not study?

d. How many students got a 90% on the test? How many hours did they study?

e. Describe the correlation, if one exists.

f. Use your graphing calculator to determine a line of best fit. Write the equation of the line of the best fit.

6. Cherie is measuring the relationship between the temperature of pond water and the number of a certain organism present in a one ounce sample. She records the data as ordered pairs. The first number in each ordered pair is the temperature of the water in degrees Celsius and the second number is the number of living organisms present.

(26, 12), (36, 6), (15, 20), (34, 8), (20, 17), (24, 18), (36, 5), (18, 22), (37, 4), (26, 11), (38, 3), (31, 10), (30, 12)

   a. Draw a scatter plot for the data.

   b. Does the data show a correlation between the temperature and the number of organisms? If so, what type of correlation?

   c. At what temperature were there the least number of organisms?

   d. Describe the relationship between temperature and the number of organisms.

   e. Enter your data into your graphing calculator and find the line of regression.

   f. What is the correlation coefficient? Explain what this means.

Mixed Review

7. Evaluate $f(x) = 5x - 15$ when x is 2, 4, and 6.

8. Evaluate $h(y) = y^2 - 2y - 12$ when y is -1, 0, and 1.

9. Evaluate $g(w) = w - 6 - \frac{w}{2}$ when w is -2, 1, and 8.

Lesson 7.4 Stem-and-Leaf Plots

Objectives

- Create a stem-and-leaf plot.
- Interpret a stem-and-leaf plot.

A **stem-and-leaf plot** allows you to quickly check the range of the data, while displaying every element of the set. This type of plot is an organized method for displaying data where the greatest and least values can quickly be identified.

The **leaves** of a stem-and-leaf plot are the set of last digits, usually the digits in the ones place. The **stems** are the digits in front of the leaves. The stems may be the digits in the tens place or they may contain several digits representing several place values. The **key** indicates the meaning of each stem and leaf.

Activity Making a Stem-and-Leaf Plot

Make a stem-and-leaf plot using the weekly hours worked data for an office temporary agency.

Weekly Hours Worked		
43	26	27
33	18	27
30	31	31
33	19	41
40	28	6
29	40	19
40	39	

Stems	Leaves
0	
1	
2	6 7
3	
4	3

Key: 2|6 = 26

1. The stems will be the digits in the tens place. Name the stems.
2. Draw a table and label the left column *Stems* and the right column *Leaves*.
3. Fill the left column with the stems. Because there are values from 0 to 4, there will be 5 rows. Draw a vertical line next to the stems.
4. Next, work your way through the data list, writing a leaf in the second column for each data value. For the first value, 43, write in a leaf of 3 beside the 4 stem. The second value gets a leaf of 6 beside the 2 stem. The third value gets a leaf of 7 beside the 2 stem, and so on.
5. Complete the plot for the rest of the values. Add a key such as 2 | 6 represents 26.
6. An ordered stem-and-leaf plot is more useful. So copy the list of stems to a new plot. Next, copy the leaves for each stem in order, from least to greatest.
7. Use your stem-and-leaf plot to find the median and mode of the hours worked data.

Notice that the stem-and-leaf plot you made in the Activity looks similar to a histogram. But, instead of tic marks or a shaded bar, your plot shows values.

Critical Thinking Why is a stem-and-leaf plot better than a histogram for finding median and mode?

Example 1 Interpreting a Stem-and-Leaf Plot

Joann's aerobic class takes a pulse rate check after 20 minutes of intense exercise. The plot at the right shows the pulse rates of the 15 participants in the class.

Pulse Rate

Stems	Leaves
10	9
11	1 5 5 8 9
12	1 4 8 8 9
13	1 2 3
14	0

Key: 12|8 = 128 bpm

a. How many people's pulse rate exceeds 120 beats per minute?

b. What is the median pulse rate for the class?

c. What is the range of the data?

Solution

a. Count the numbers in the stem rows of 12, 13, and 14. There are nine participants with pulse rates over 120 beats per minute.

b. Because the class has 15 people (an odd number), the median is an entry in the plot. The median is the 8th entry or 124 bpm.

c. The lowest pulse rate is 109 bpm and the highest is 140 bpm. So the range of the data is 31 bpm (140 − 109).

Example 2 Batting Averages in a Stem-and-Leaf Plot

Coach Yablonsky calculated the batting average of the players on his team after 30 games. The plot at the right shows the batting average of the 20 players on this team.

Batting Averages

Stems	Leaves
0.29	3 4 5 9
0.30	0 1 8
0.31	6 7
0.32	8
0.33	4 5 5 6
0.34	1 7 9
0.35	4 4 7

Key: 0.31|6 = 0.316

a. What is the lowest batting average on this team? the highest?

b. What is the range of batting averages?

c. What percentage of hitters has averages of 0.335 or greater?

d. What is the median batting average for this team?

e. What is the mode of these batting averages?

Solution

a. The lowest batting average is shown as 0.29 | 3, or 0.293. The highest is 0.357.

b. The range is $0.357 - 0.293 = 0.064$.

c. In the 0.33 stem, you see that there are three values of 0.335 or greater. In the 0.34 stem there are three values, as also in the 0.35 stem. So there are 9 batters at 0.335 or greater. Counting the leaves, there are 20 batters. So $\frac{9}{20}$, which is 45%, have averages of 0.335 or greater.

d. Because there is an even number of leaves, the median is the average of the two middle values: 0.32 | 8 and 0.33 | 4.

$$\text{median batting average} = \frac{0.328 + 0.334}{2} = 0.331$$

e. There are two values that occurs the same number of times. The mode is 0.335 and 0.354.

Lesson Assessment

Think and Discuss

1. Name at least one advantage to organizing data in a stem-and-leaf plot.

2. Explain how to determine what digits are the stems and what digits are the leaves.

3. Name the stems of the following set.
24 35 46 14 15 55

Practice and Problem Solving

4. Draw a stem-and-leaf plot for these quiz scores on a recent math quiz: 71, 88, 86, 66, 76, 80, 69, 76, 86, 79, 59, 90, 91, and 58.

5. Find the median and mode of the test scores in Exercise 4.

6. Draw a stem-and-leaf plot for the grade-point average of ten students applying for a scholarship: 3.85, 3.75, 3.66, 3.96, 3.68, 3.76, 3.89, 3.95, 3.87, and 3.75.

7. Find the median and mode of the grade-point averages in Exercise 6.

8. Use the landscape company data shown in Lesson 7.2 on page 402.

a. Make an ordered stem-and-leaf plot.

b. Compare this plot with the histogram you drew in Lesson 7.2. Which presents more information? Explain.

c. On how many days did the company trim more than 20 trees?

City High/Low Temperatures on October 1

High		Low
	2	5
8	3	2 8
5 7	4	4 7 9
8 0	5	2 5 9
2	6	
9	7	1 4 4
1 5 9	8	0

Key: |8|0 = 80
: 9|8| = 89

9. Use the double stem-and-leaf plot shown at the left.

a. How many cities had low temperatures below 40 for October 1?

b. How many cities had high temperatures above 80 for October 1?

c. Which has the greater range, the low temperatures or the high temperatures?

d. How many times did a city's high temperature equal another city's low temperature on October 1?

Mixed Review

Find the slope of each line. Name the slope of a line parallel to the given line. Name the slope of a line perpendicular to the given line.

10. $2x + 4y = 1$

11. $y = -3$

12. $5x - y = 5$

13. A rectangular prism that has a height of $4x$, a width of $2x + 1$, and length of $\frac{x}{2}$.

a. Write the expression for the volume of the prism.

b. Write the expression of the surface area of the prism.

Lesson 7.5

Quartiles and Box-and-Whisker Plots

Objectives

- Calculate quartile values for a data set.
- Create a box-and-whisker plot.

Richard recorded the number of people that attended his team's minor league hockey games this year. Identify the first, second, and third quartiles of his data.

Game Attendance			
211	246	242	305
195	272	296	140
279	250	246	265
242	209	311	

Quartiles

Quartiles are boundaries that break the data into fourths. You already know how to find the second quartile, Q_2, because the second quartile is better known as the median.

The first quartile, Q_1, is the median of the lower half of the data. The third quartile, Q_3, is the median of the upper half of the data.

One-fourth of the data in a set falls at or below Q_1. One-fourth of the data falls between Q_1 and Q_2. One-fourth of the data falls between Q_2 and Q_3. One-fourth of the data falls at or above Q_3.

Example 1 Finding Quartiles

Use Richard's data from the opening paragraph of this lesson. Identify the first, second, and third quartiles.

Solution

First sort the data from least to greatest. Then locate the median. The set has 15 numbers so the median is one of the numbers in the set. The median, or second quartile, is 246.

Because there are seven numbers in the lower half of the data, the first quartile is 211.

Because there are seven numbers in the upper half of the data, the third quartile is 279.

$$140\ \ 195\ \ 209\ \ \underset{Q_1}{\underset{\uparrow}{211}}\ \ 242\ \ 242\ \ 246\ \ \underset{\text{median, or } Q_2}{\underset{\uparrow}{246}}\ \ 250\ \ 265\ \ 272\ \ \underset{Q_3}{\underset{\uparrow}{279}}\ \ 296\ \ 305\ \ 311$$

Box-and-Whisker Plots

A **box-and-whisker plot** is a visual presentation of the data that divides the data into quartiles, or fourths. This plot also shows the quartile values, and the least and greatest numbers in the set.

One advantage of displaying data in a box-and-whisker plot is that it does not become more difficult to interpret the plot as the number of data values increase. One disadvantage of displaying data in a box-and-whisker plot occurs when there are not many data values.

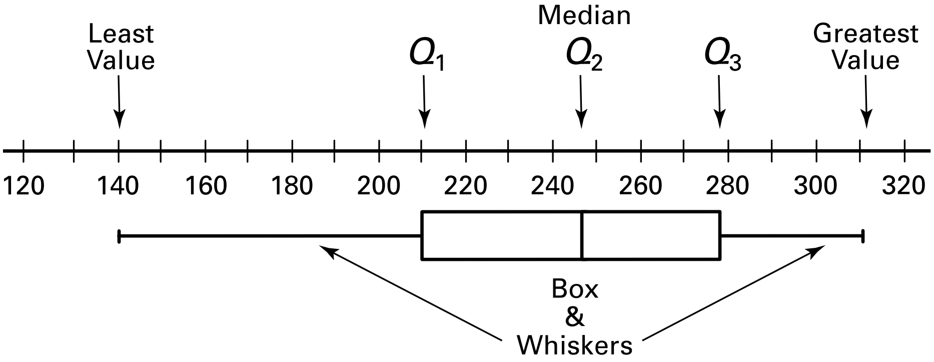
Example 2 Creating a Box-and-Whisker Plot

Use Richard's data from the opening paragraph of this lesson to create a box-and-whisker plot.

Solution

Draw a number line where you can easily locate the least value, the greatest value, and each of the quartile values.

A box is drawn below the number line that is the length from Q_1 to Q_3. The box has a vertical line drawn to separate it into two parts at Q_2. Points are made at the least value and greatest value. Then lines, known as whiskers, are drawn to connect these points to the ends of the box.



If an *outlier* is part of the set, it is shown but noted as an outlier. An outlier is a data value that is clearly not near (or close in range) to the rest of the data. Outliers are unusually small or unusually large values.

The interquartile range is $Q_3 - Q_1$. You can usually conclude that a data value is an outlier if it lies more than 1.5 times the interquartile range from Q_1 or Q_3.

Activity Making a Box-and-Whisker Plot

Make a box-and-whisker plot using the weekly-hours data from an earlier Activity.

Weekly Hours Worked Office Temp, Inc.		
43	26	27
33	18	27
30	31	31
33	19	41
40	28	6
29	40	19
40	39	

1 First, sort the data from least to greatest.

2 Find the median value, Q_2.

3 Find the first-quartile value, Q_1.

4 Find the third-quartile value, Q_3.

5 Identify the least and greatest values—the whisker values.

6 Draw a scale for your plot that will include the whiskers.

7 Locate the values for Q_1, Q_2, and Q_3 on the scale. Draw the box spanning these three values.

8 Locate the least and greatest values on the scale. Draw the whiskers from the box ends to the least and greatest values.

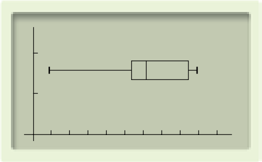
Note that when calculating the medians of the lower or upper quartiles of a data set that has an odd number of data values does not include the overall median in its calculation.

Example 3 Using Technology

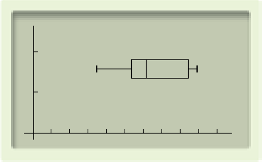
a. Use the data from the Activity and your graphing calculator to verify the box-and-whisker plot you created above.

b. Determine how to adjust the list in your graphing calculator to reflect that the data set has an outlier.

Solution



a. To enter the data into a list, press STAT. Select Edit. Enter the data into L1. Press 2nd STAT PLOT and select Plot1. Turn on Plot 1 and select the box-and-whisker plot for type. Be sure that the Xlist named is L1. Set your viewing window to show the first quadrant of the coordinate plane. Because the Yscl is not used for the box-and-whisker plot set the minimum at -1 and maximum at 2. To view your graph, press GRAPH.



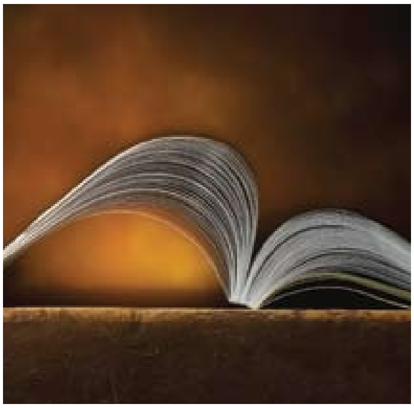
b. Because you want to use 18 as one whisker, delete 6 from the L1 list.

The outlier can be plotted using Plot2 and the L2 and L3 lists.

CULTURAL CONNECTION

The definition of *quartile* is as follows:

quartile—*n* [ISV, fr. L *quartus*] (1879): any one of three values that divides the items of a frequency distribution into four classes with each containing one fourth of the total population: *also* any one of the four classes



The information within brackets that precedes the definition is known as the *etymology* of the word. An etymology traces the history of a word in English and tells what language it came from and in what form it is used in English.

Many words are traced to either Middle English (ME) or Old English (OE), or both. In the case of mathematical terms, many have a label of ISV which stands for International Scientific Vocabulary. Such a reference means the etymology recognizes the use of this word as international and the possibility exists that it originated somewhere other than the English language.

The fr. L quartus states that the word comes from the Latin word *quartus* meaning fourth. When a word is changed slightly for the English language, the original word in the original language is called a *transliteration*.

1. Check the etymology of *quartile* in a different dictionary to see if more detail of the history of the word is given.
2. Choose another mathematical term and investigate its etymology.

Lesson Assessment

Think and Discuss

1. Can you find the mean of the data from a box-and-whisker plot? Explain.
2. Excluding any outliers, how can you find the range of the data from just looking at a box-and-whisker plot?
3. Describe how to find each of the quartiles.

Practice and Problem Solving

Car Prices

11,300	11,600	11,400
10,500	11,600	10,500
11,600	12,000	11,600
11,400	11,700	11,800
11,000	11,800	12,100

4. Ella plans to sell her sports car. On the Internet she finds many prices from other people trying to sell cars of the same kind and age. Ella made a list of the prices, as shown, rounded to the nearest $100.
   - **a.** What is the Q_2 value?
   - **b.** What is the Q_1 value?
   - **c.** What is the Q_3 value?
   - **d.** Name the whiskers.
   - **e.** Construct a box-and-whisker plot of these values.
   - **f.** After seeing your plot, do you feel that you might have an outlier in your data? If so, remove the outlier and use your graphing calculator to redo your plot.
   - **g.** Which plot, with or without the outlier, is a better guide for Ella to use to determine the price at which she should sell her car?
   - **h.** Find the quartile values and the mean of the data excluding the outlier. What affect does excluding the outlier have on each value?

Mixed Review

Name the property illustrated in each statement. Determine the value of *x*.

5. $2x = 0$

6. $\frac{6}{x}\left(\frac{x}{6}\right) = 1$

7. $-4(x + 3) = -20 - 12 = 32$

8. $5x + 6 + 7x = 12x + 6 = 30$

Math Labs

Activity 3: Using Technology to Create Statistical Graphs

Problem Statement

You will collect data about the different ways teenagers spend their days.

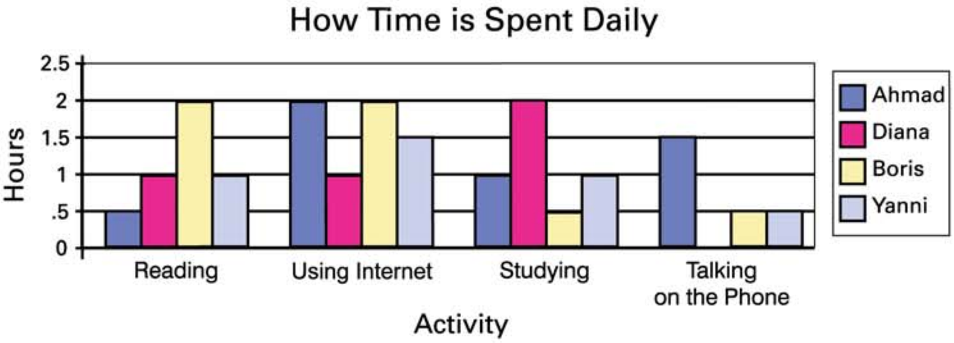
Collect data from four of your classmates. Enter the data into a spreadsheet to create a table. Then make at least two different types of graphs for the data.

Equipment

Spreadsheet software application

Procedure

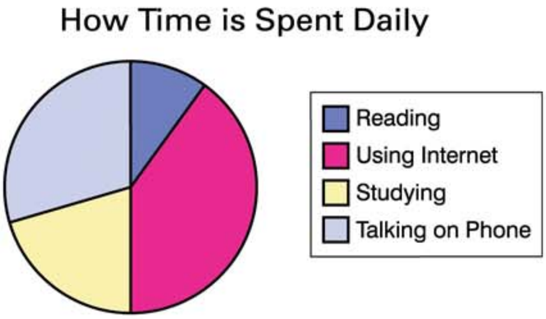
1. List four ways you pass your time while at home, such as watching TV or reading.
2. Ask four classmates how many hours they spend each day on each of the activities named in Step 1.
3. Open a new document in the spreadsheet program. Name your file.
4. Set up five columns, one for each activity and one column for students' names. Make four rows, one for each student.
5. Enter the data into the appropriate cells in your file.
6. Use the spreadsheet program to create a bar graph. In the *Insert* menu, select *Object*. You will see illustrations of the different types of graph. Click on the type of bar graph you want to create.



7. Follow the prompts to title your graph and axes. A sample bar graph is shown above.

8 Use the spreadsheet to create a circle graph. In the *Insert* menu, select *Object*. Click on the type of graph you want to create.

9 Follow the prompts to title your graph and axes. A sample circle graph is shown.



10 Compare the two different graphs you created. Even though each was made with the same data, they look quite different. By looking at the graphs only, do they show the same information? Explain.

11 Write a question that can be answered by looking at the bar graph.

12 Write a question that can be answered by looking at the circle graph.

13 Which graph displays all the individual data in the table?

14 Which graph displays how parts of the data relate to the whole?

15 Which graph can you use to quickly rank the four activities in the order of most time spent to least time spent for your four classmates? What is the ranking?

16 Describe how each graph could change if you increased the number of rows in the table, and listed all the students in your class with their responses.